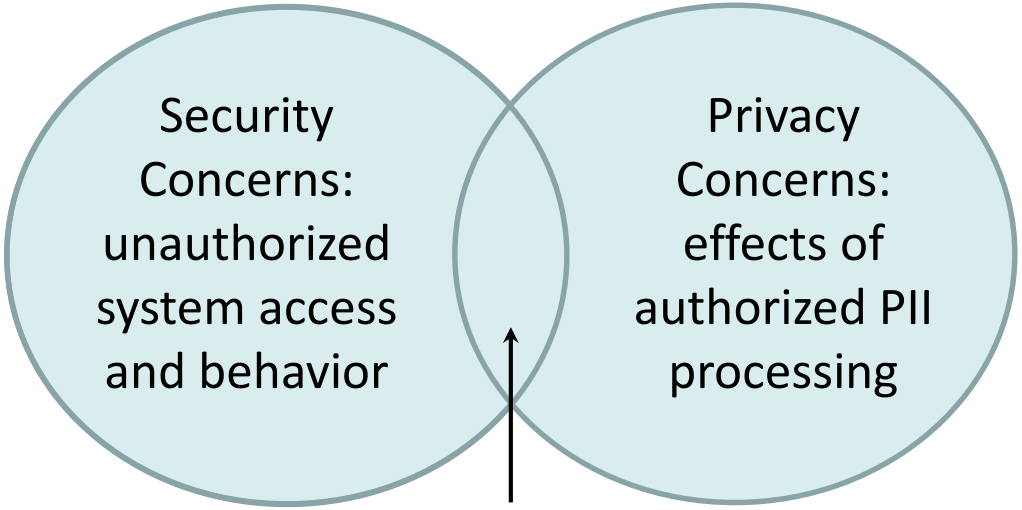
Networking and Information Technology Research and Development Program

ACSAC 2017
NITRD Panel: Big Data for Security
Can We Improve Security and Preserve Privacy?

Joshua Baron	Program Manager, DARPA
Jeremy Epstein	Deputy Division Director, NSF
Steven King	Deputy Director, Cyber Technology, Office of the Assistant Secretary of Defense
Scott Tousley	Deputy Director, Cyber Security Division, DHS S&T
Tomas Vagoun	R&D Coordinator, NITRD

On Information Security and Privacy

Challenge: build systems that satisfy technical requirements



Challenge: build systems that satisfy social requirements: privacy expectations (norms and laws)

Security of PII

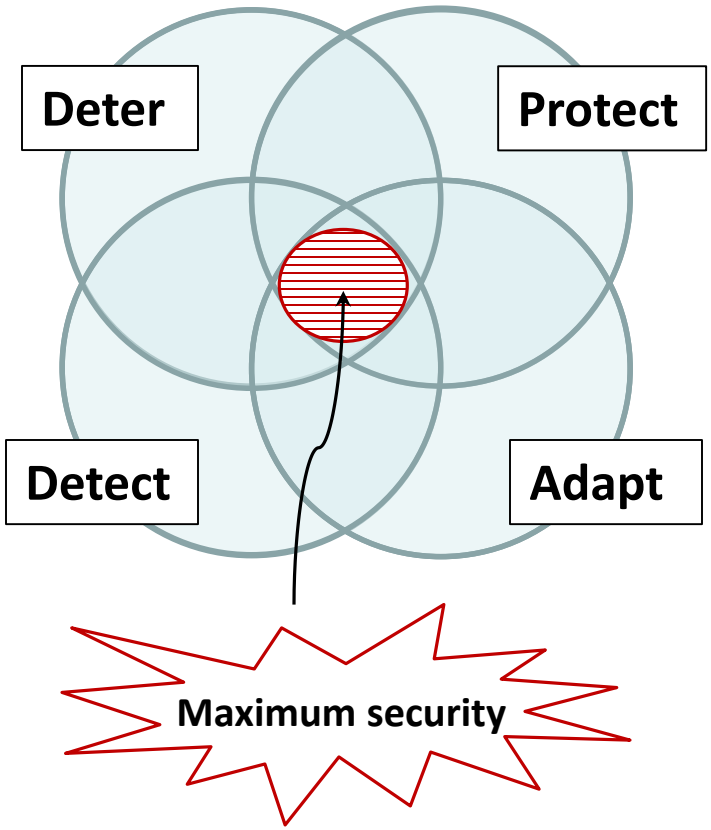
<u>Security Engineering Objectives</u>

- Confidentiality
- Integrity
- Availability
- Nonrepudiation
- …

<u>Privacy Engineering Objectives</u>

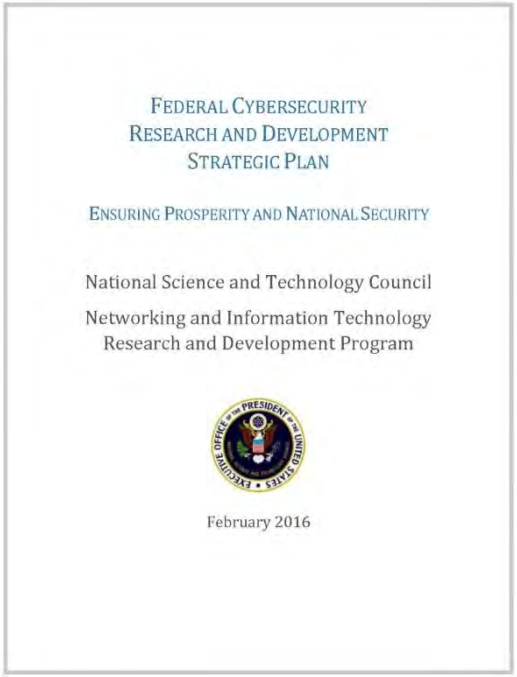
- Predictability (contextual integrity)
- Disassociability (unlinkability)
- Manageability (intervenability)
- Transparency
- …
- [see NIST IR 8062/Privacy Engineering]

Strategic Plan for Federal Cybersecurity R&D



Federal Cybersecurity R&D Goals

- S&T for **effective and efficient risk management**
- S&T for **sustainably secure systems development and operation**
- S&T for **effective and efficient defensive deterrence**



FEDERAL CYBERSECURITY
RESEARCH AND DEVELOPMENT
STRATEGIC PLAN

ENSURING PROSPERITY AND NATIONAL SECURITY

National Science and Technology Council

Networking and Information Technology
Research and Development Program

February 2016

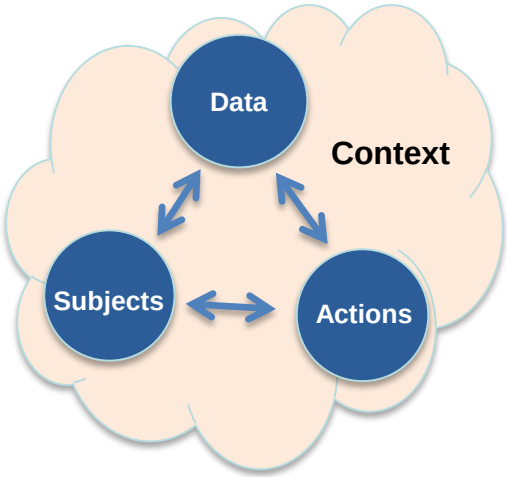
Critical Dependencies

- Scientific foundations
- Risk management
- Human aspects
- Transition to practice
- Workforce development
- Infrastructure for research

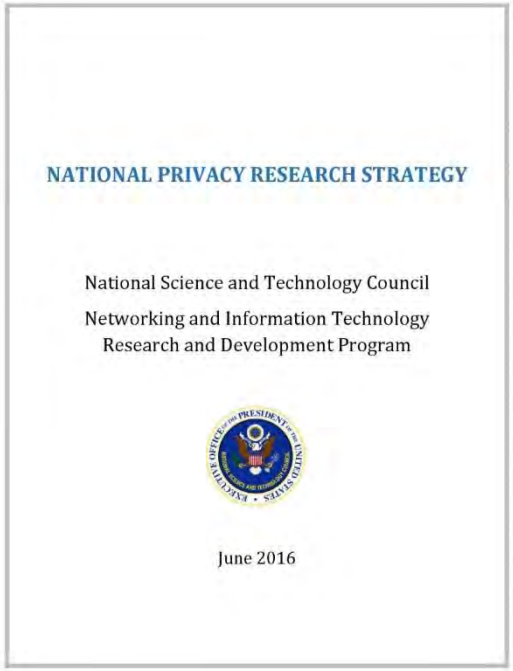
NITRD

Strategic Plan for Federal Privacy R&D

Privacy As



- Foster multidisciplinary approach to privacy research and solutions
- Understand and measure privacy desires and impacts
- Develop system design methods that incorporate privacy desires, requirements, and controls
- Increase transparency of data collection, sharing, use, and retention
- Assure that information flows and use are consistent with privacy rules
- Develop approaches for remediation and recovery
- Reduce privacy risks of analytical algorithms


NATIONAL PRIVACY RESEARCH STRATEGY

National Science and Technology Council

Networking and Information Technology Research and Development Program

June 2016

NITRD

Strategic Plan for Federal Big Data R&D

Next Generation Capabilities

- New models, architectures, statistical methods, networks, algorithms

Cyberinfrastructure

- Build and sustain research cyberinfrastructure for BD innovation

Privacy & Ethics

- Provide equitable privacy protections
- Understand ethics for data governance
- Enable secure BD cyberspace


THE FEDERAL BIG DATA RESEARCH AND DEVELOPMENT STRATEGIC PLAN

THE NETWORKING AND INFORMATION TECHNOLOGY RESEARCH AND DEVELOPMENT PROGRAM

NITRD

MAY 2016

Education

- Expand the pool of data scientists
- Broaden data-capable workforce
- Expand the community of data-empowered domain experts

Partnerships

- Encourage cross-sector, cross-agency BD collaborations
- Promote policies/frameworks for faster responses

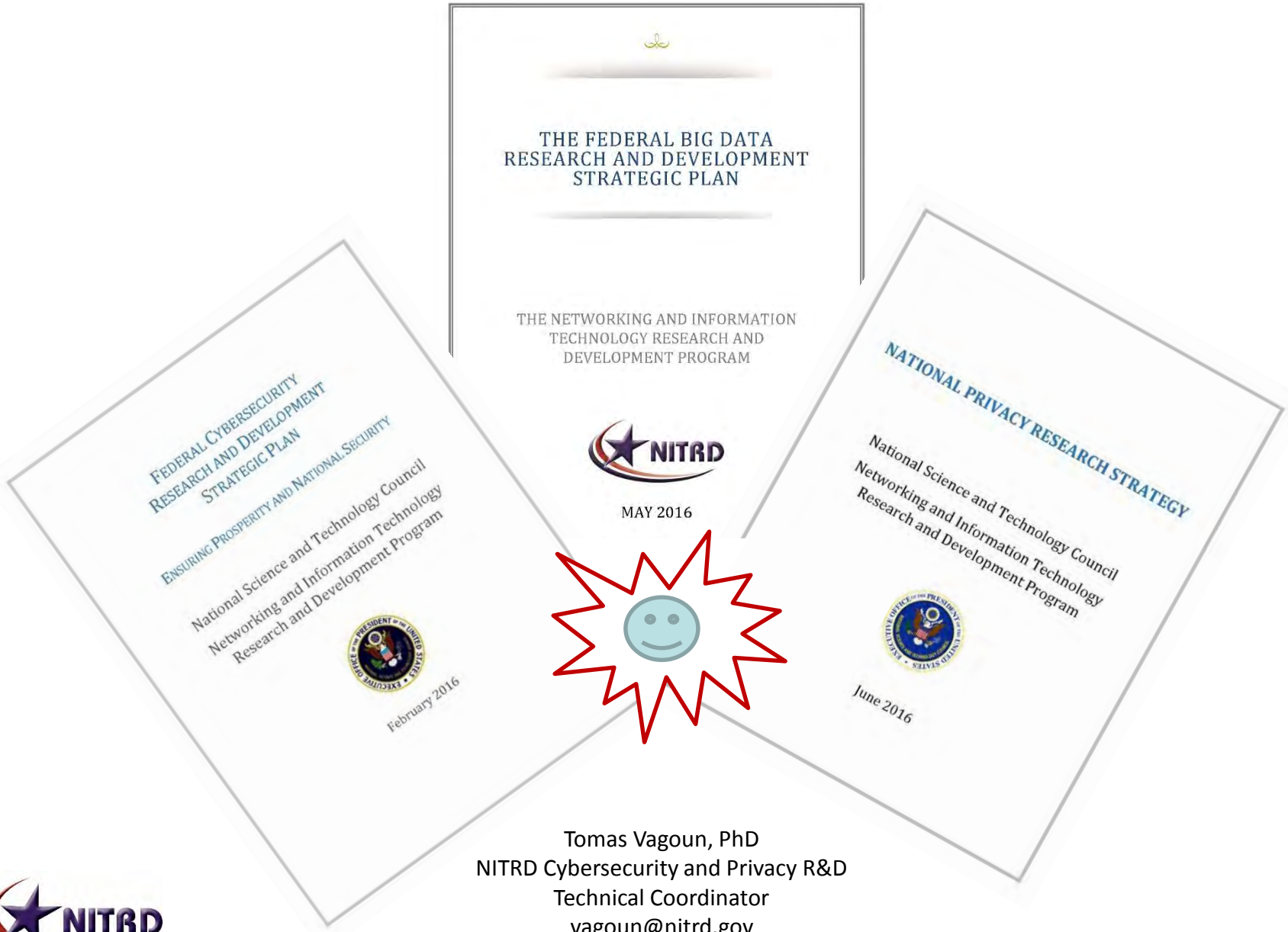
Data Sharing

- New metadata standards, frameworks, APIs, ontologies

Trustworthiness & Decision-making

- New statistical tests, metadata management, curation, multi objective data driven support systems

NITRD



Tomas Vagoun, PhD
NITRD Cybersecurity and Privacy R&D
Technical Coordinator
vagoun@nitrd.gov

Brandeis Program

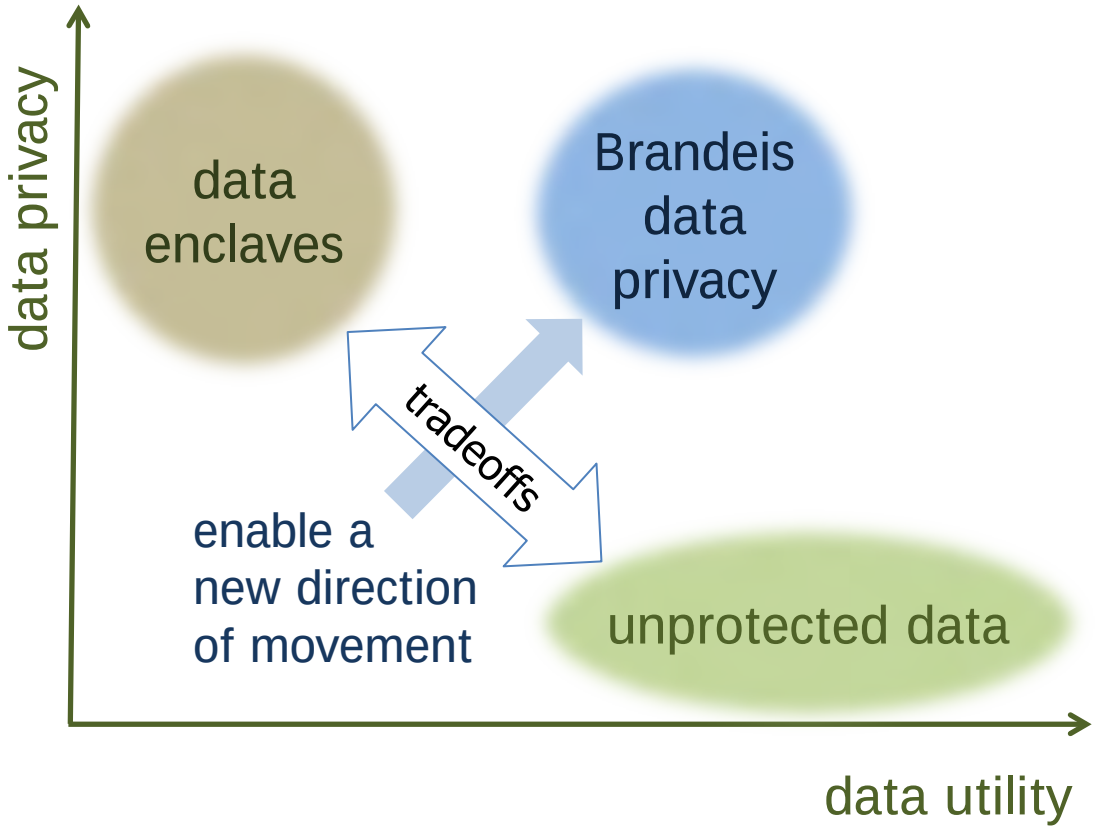
Dr Josh Baron
Program Manager

ACSAC big data privacy panel

12/06/2017

Program Objective

Develop tools and techniques to enable the building of information systems where private data can be used for the intended purpose – and no other

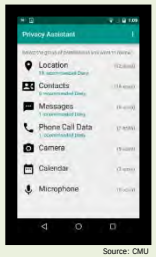


Brandeis System Concept

Data Owner

& Devices

Human Data Interaction (TA2)



Source: CMU

Privacy Policy User Interface

(Users define and understand policy)

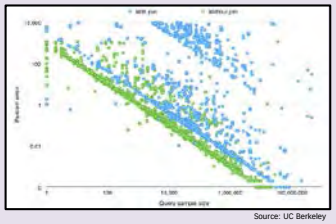
Experimental Integration Platforms (TA3)

TA1s TA1s TA1s TA1s

Privacy Preserving Techniques

(Encryption, MPC, Differential Privacy)

Data Requester / User / Analyst



Source: UC Berkeley

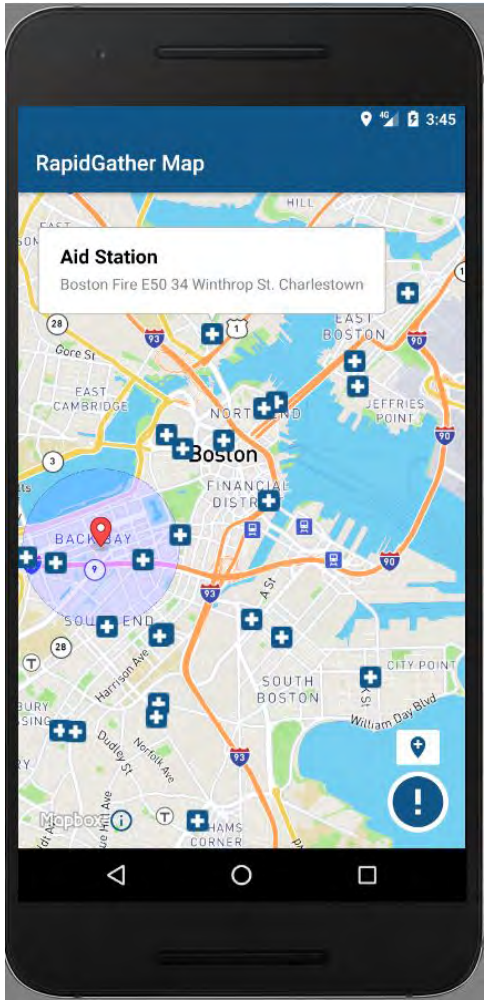
Data requesters use, analyze privacy protected data

Privacy Measurement (TA4)

(Metrics for measuring privacy protection)

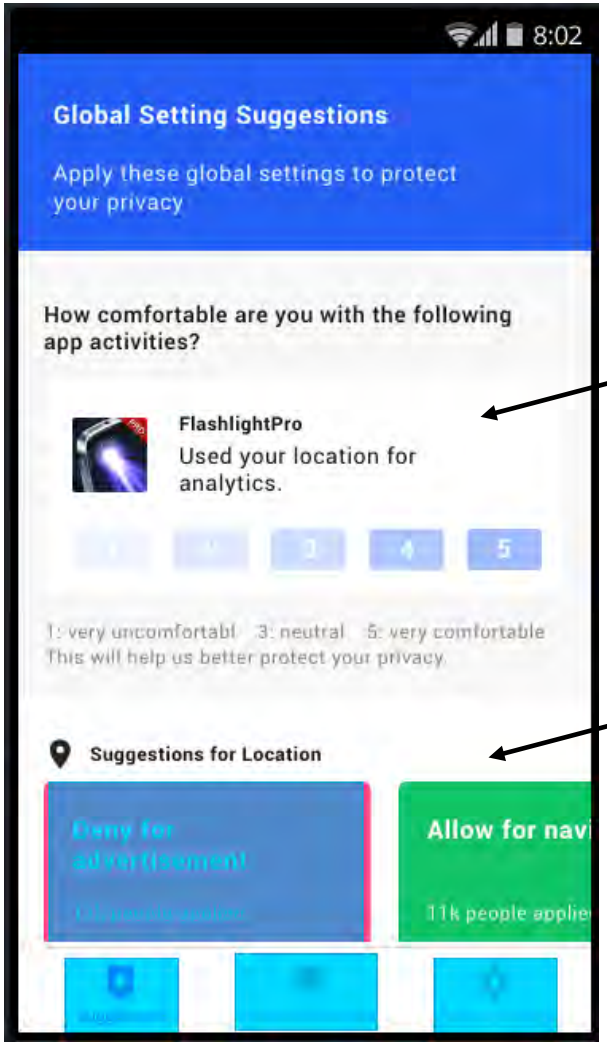
Privacy-Enhanced (PE) Android

PE Android with RapidGather App



Source: Raytheon BBN

Prototype <u>simple</u> interface for privacy decisions



Infer user privacy prefs

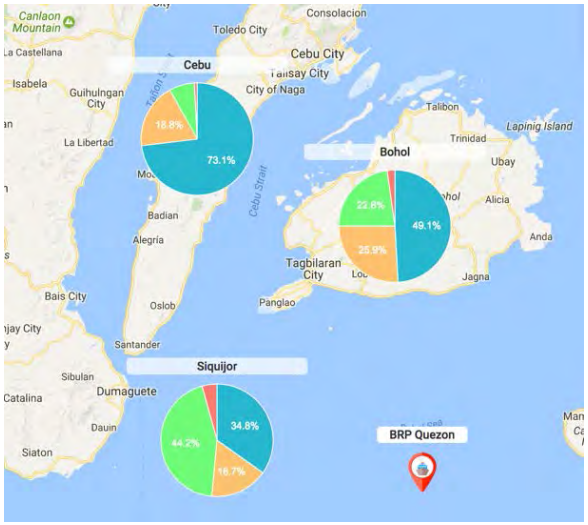
Crowdsource privacy reasoning

Source: CMU

Coalition Information Sharing

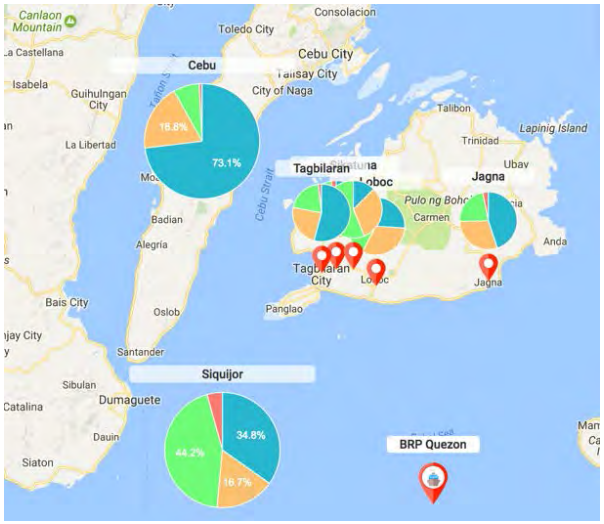
Policy-differentiated Information Access:

- Different content and resolution based on privacy policies for different community, national, and international roles



International Response Coordinator: *nation-level view, but no community level details*

Cebu City Coordinator: *community-level view only*



Source: SRI

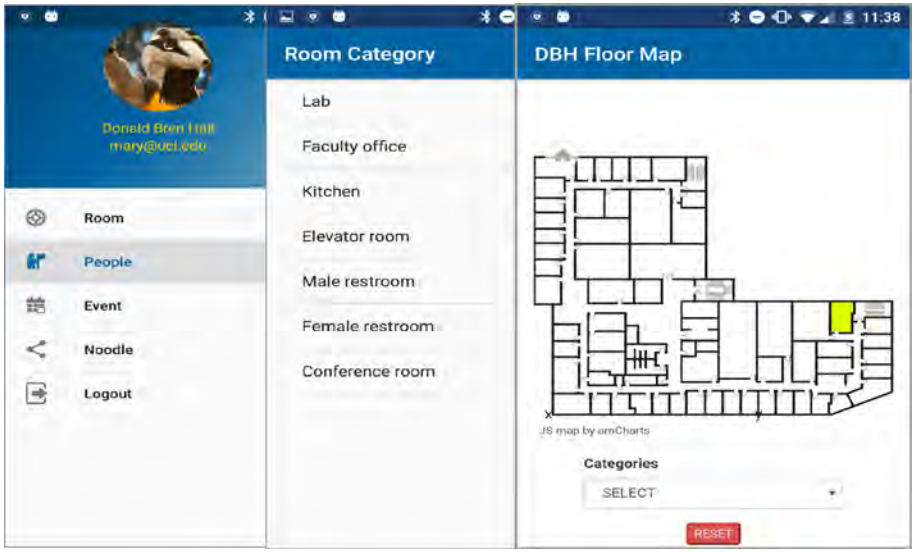
Bohol Coordinator: *community-level view for Bohol, nation-level view for others*

UC Irvine smart building privacy testbed

Building testbed for researching and experimenting with privacy technologies:

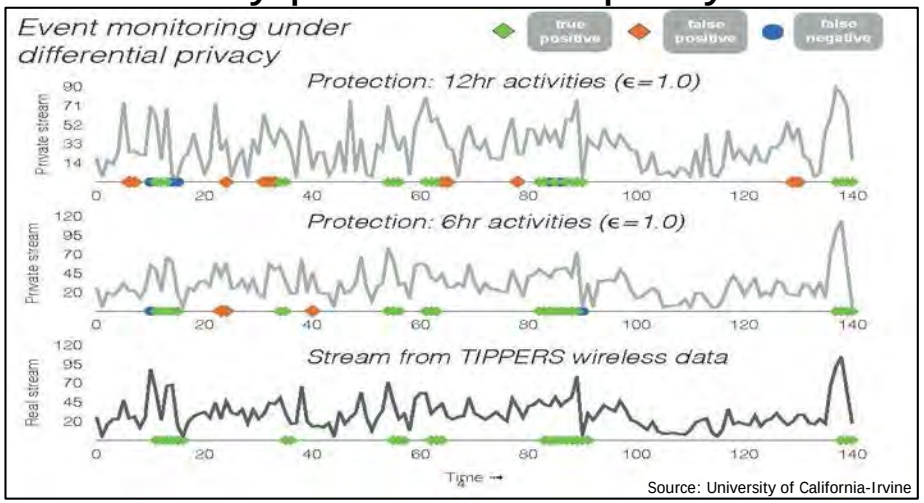
- Physical and virtual sensor feeds
- API for rapid creation of privacy preserving apps that integrate with building sensor data
- Integration of diverse privacy-preserving compute & storage technologies

Building concierge



Source: University of California-Irvine

Privacy-protected occupancy data



Source: University of California-Irvine

www.darpa.mil

Security, Big Data & Algorithmic Accountability

The NSF Secure & Trustworthy Computing Program

Jeremy Epstein, Deputy Division Director
Computer and Network Systems

NSF research programs in Big Data . . .

Encompass Research, Cyberinfrastructure, Education and Training, and Community Building

Cover algorithmic, statistical, and mathematical foundations of data science; new techniques, technologies, and methodologies, including hardware and software approaches; and innovative uses of data for scientific discovery and action

Similar to SaTC, Big Data projects cut across divisions and directorates at NSF

Big Data & SaTC

Big Data is transforming research in all areas of science and engineering, cybersecurity included

The accumulation of large amounts of personal data by government, companies, and other organizations has important implication for the privacy and security of personal data

Recent projects & workshops at the intersection of SaTC and Big Data

- Secure Data-Intensive Computing on Hybrid Clouds
- Privacy Preserving Computation in Big Data Clouds
- Workshop: Advancing ethics for trustworthy cyberspace and data analytics

Secure Data-Intensive Computing on Hybrid Clouds

Award Number: 1223495
Indiana University;
XiaoFeng Wang & Geoffrey Fox
2012; Small; $500,000 for 5 years

Data-intensive computations traditionally has been done on individual organizations' internal systems due to concerns with low-cost public clouds adequately protecting sensitive user data

For cloud-based solutions to be practical, privacy concerns must be addressed

Challenge: existing cryptographic techniques tend to be too heavy-weight to manage large amounts of data

Solution: develop privacy-aware MapReduce system that partitions components across public/private clouds according to security levels required by data, in way that is efficient and secure

Project involves industry collaborators, and advances may be useful for wide range of computing jobs (e.g., commercial data analysis; DNA analysis; intrusion detection)

Privacy Preserving Computation in Big Data Clouds

Award Number: 1564097
Georgia Tech Research Corporation
Ling Liu & Calton Pu
2016 Medium; $1,199,999 for 5 years

Privacy is vital to freedom of creativity and innovation, and must be protected if we are to achieve maximum benefits from harnessing big data

For cloud-based solutions to be practical, privacy concerns must be addressed

Challenge: the ability to perform efficient big data computations in the cloud has great potential for data analytics related to health, advertising, and other domains, but there are many concerns with user privacy

Solution: The PrivacyGuard project is a practical framework that seeks to enhance privacy-preserving distributed computation by creating algorithms, systems, and tools that guarantee end-to-end privacy throughout a data analytic job

- Designing formal mechanisms for privacy requirements for data release (e.g., associating data release with usage framework to restrict the analyses that may operate)

- Developing set of guards intended to audit and enforce compliance during analysis
- Devising proactive strategy to prevent information leakages associated with mining output

Integration of research with curriculum development of Georgia Institute of Technology contributes to ensuring future data scientists are aware of privacy

Workshop: Advancing ethics for trustworthy cyberspace and data analytics

Award Number: 1623445
Virginia Tech
Susan Sterett & Kelly Joyce
2016 Workshop; $47,455

Big data analytics centers have the potential to address a broad set of problems (e.g., health disparities, natural disasters; social stability in urban settings)
At the heart of many of these problems are central ethical questions related to privacy, inequalities, validity, and use

Challenge: Collaboration between users and developers of new analytic tools is important for creating meaningful ethical practices; however, a disconnect exists between developers and users

Solution: Hold a workshop that brings together a diverse set of stakeholders with different perspectives, (e.g., developers; ethnographers of scientific practices; users), with goal of building framework for big data creation that emphasizes consideration of ethical issues

USACM Seven Principles for Algorithmic Transparency and Accountability

1. Awareness
2. Access and redress
3. Accountability
4. Explanation
5. Data Provenance
6. Auditability
7. Validation and Testing

https://www.acm.org/articles/bulletins/2017/january/usacm-statement-algorithmic-accounta

Questions?

Jeremy Epstein
Deputy Division Director, Computer & Network Systems
Directorate of Computer & Information Science & Engineering
jepstein@nsf.gov
703-292-8338

Big Data & Data Analytics for Security

Dr. Steven E. King
Deputy Director, Cyber Technology
Office of the Assistance Secretary of Defense
(Research & Engineering)

Approved for public release; OSD Case # 15-S-1708

Data Analytics Assessment Overview

Problem: There is currently no standard way to implement and assess performance for data analytics

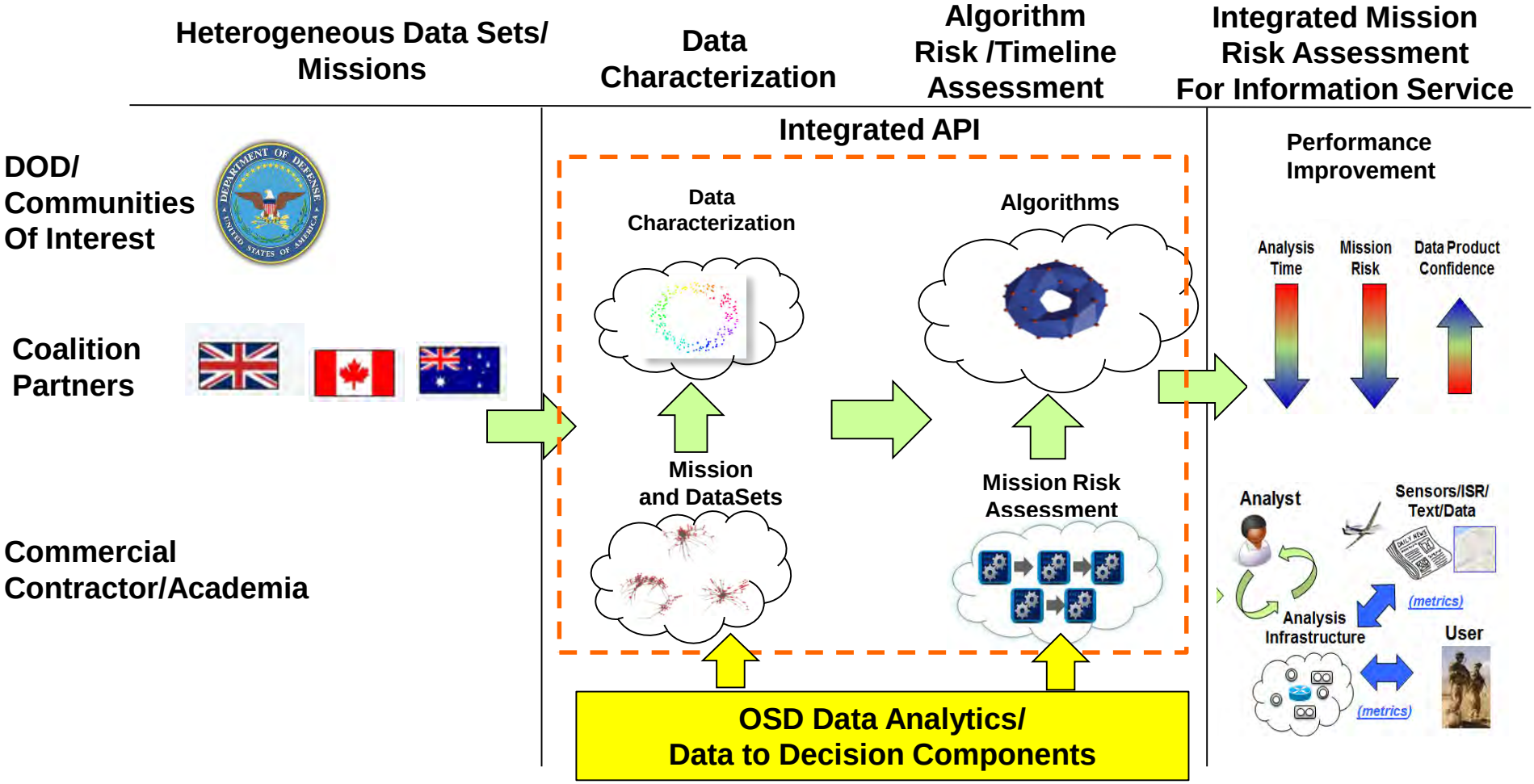
- Heterogeneous data sources/algorithms without ground truth
- Hard to know what capability is being purchased with few means to assess performance of service
- Dynamic mission space with changing requirements

Solution: Data analytics framework

- Standard data models with ground truth
- Development framework to standardize risk analytics on information sources, algorithms, and processing
- Adaptable framework that can change as mission requirements change

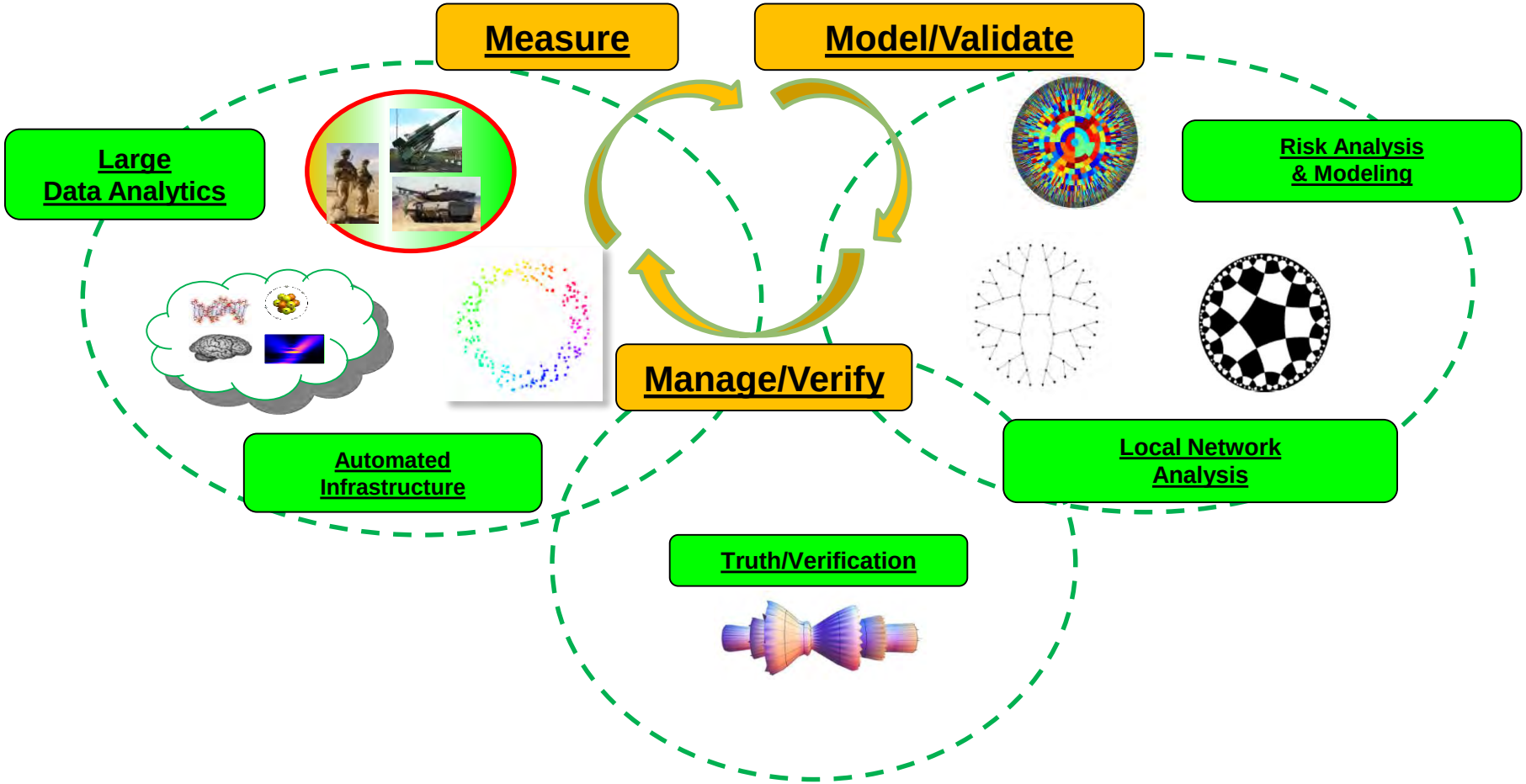
Data Analytics Performance Assessment

Implementation and assessment of information service can be standardized to assess overall mission performance



Measure/Model/Manage

Integrated modeling, validation, verification, and management can characterize mission performance with advanced data models



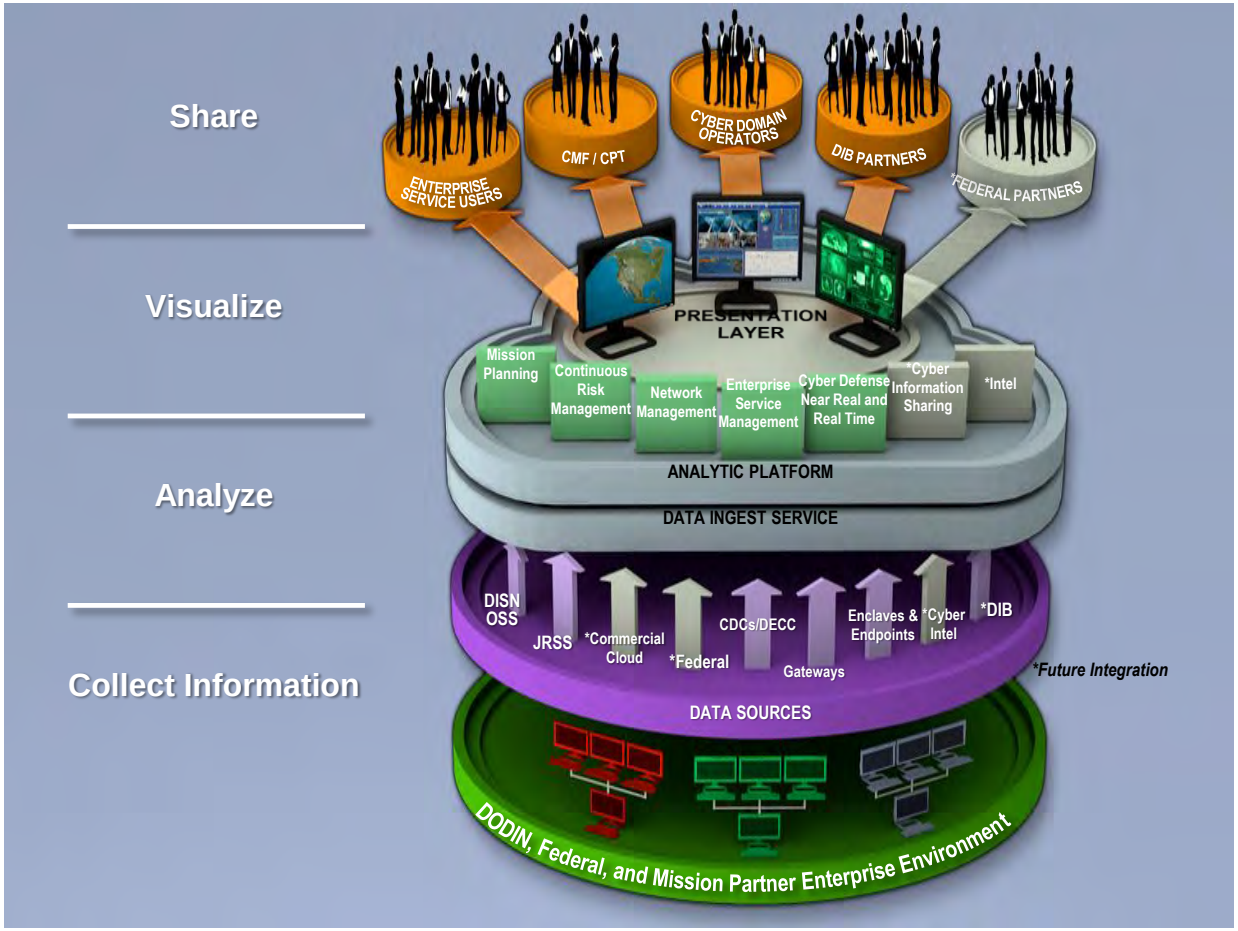
Cyber Situational Awareness (SA) and Big Data

Transforming Cyber SA with Big Data

- Enables situational awareness across Defensive Cyber Operations (DCO) and DoDIN Operations domains
- Correlates across the *perimeter*, *regional*, & *endpoint* data sources to include *threat intelligence*
- Common platform for shared analytics
- Complements real-time incident and event management for more complete picture

Quick Facts

- 3 operational instances – NIPR, SIPR, & Private Secret Enclave
- 1 year ATO with conditions under RMF
- Many existing production analytics
- 24 different data sets ingested
- 1680+ active users across all DoD services
- Over 930 collaborative developers



Cyber Situational Awareness Alignment to Cyber Strategy

Classified Reporting Data (Sept 2016)

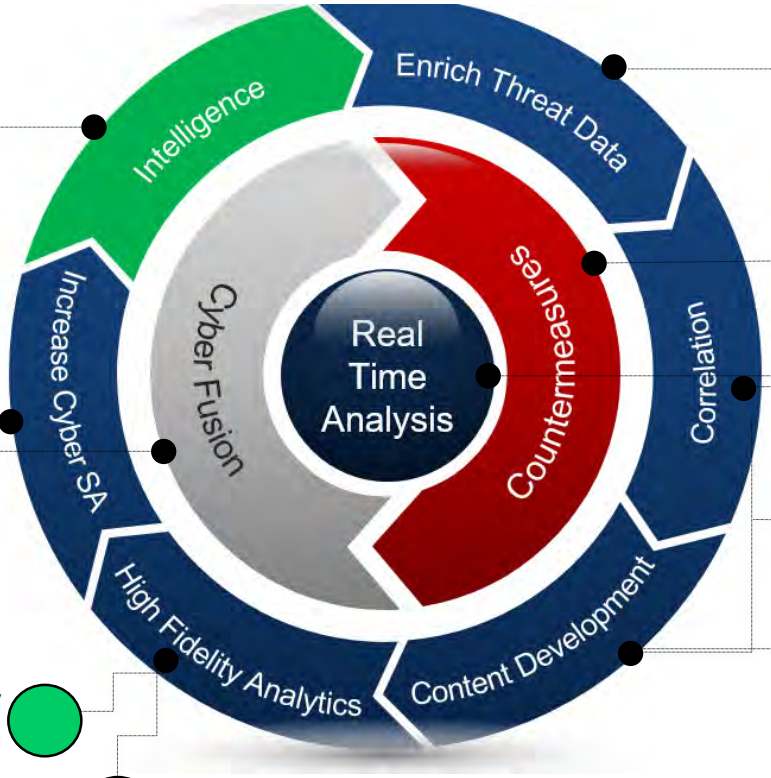
Open Source & CMCL Feeds (Q3FY18)

ACT (Dec 2017)

Unity Search (Oct 2017)

SIEM & Big Data Interoperability (Q2FY18)

Pilot Commercial Big Data (Sept 2017)



Cross Domain Services (Q3FY19)

Data Brokering (Q1FY19)

Signature Repository (Jul 2017)

Next Gen SIEM (Q3FY19)

Splunk Pilot (Sep 2017)

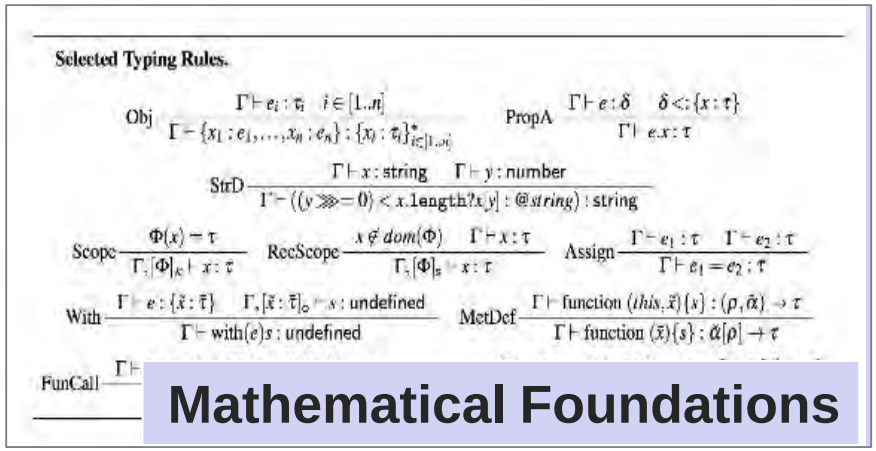
Analytic Sandbox/DEVOPS (Q2FY18)

Operational | On Glide Path | Moderate Risk | Not Started

DISTRIBUTION STATEMENT A. Approved for public release; OSD Case # 15-S-1708

Cyber S&T Vision Research Areas

Strong foundations and disruptive innovations that create surprise, shape the fight, and ensure decisive advantage



"Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Networking and Information Technology Research and Development Program."

The Networking and Information Technology Research and Development (NITRD) Program

Mailing Address: NCO/NITRD, 2415 Eisenhower Avenue, Alexandria, VA 22314

Physical Address: 490 L'Enfant Plaza SW, Suite 8001, Washington, DC 20024, USA Tel: 202-459-9674, Fax: 202-459-9673, Email: nco@nitrd.gov, Website: https://www.nitrd.gov